

An Approximation Approach for Solving the Subpath Planning Problem

Masoud Safilian¹, S. Mehdi Tashakkori Hashemi¹, Sepehr Eghbali², Aliakbar Safilian³

¹ Amirkabir University of Technology, Iran
ma.safilian@gmail.com, hashemi@aut.ac.ir

² University of Waterloo, Canada
sepehr.eghbali@uwaterloo.ca

³ McMaster University, Canada
safiliaa@mcmaster.ca

Abstract

The subpath planning problem is a branch of the path planning problem, which has widespread applications in automated manufacturing process as well as vehicle and robot navigation. This problem is to find the shortest path or tour subject for travelling a set of given subpaths. The current approaches for dealing with the subpath planning problem are all based on meta-heuristic approaches. It is well-known that meta-heuristic based approaches have several deficiencies. To address them, we propose a novel approximation algorithm in the $O(n^3)$ time complexity class, which guarantees to solve any subpath planning problem instance with the fixed ratio bound of 2. Beside the formal proofs of the claims, our empirical evaluation shows that our approximation method acts much better than a state-of-the-art method, both in result and execution time.

Note to Practitioners—In some real world applications such as robot and vehicle navigation in structured and industrial environments as well as some of the manufacturing processes such as electronic printing and polishing, it is required for the agent to travel a set of predefined paths. Automating this process includes three steps: 1) capturing the environment of the actual problem and formulating it as a subpath planning problem; 2) solving subpath planning problem to find the near optimal path or tour; 3) command the robot to follow the output. The most challenging phase is the second one that this paper tries to tackle it. To design an effective automation for the aforementioned applications, it is essential to make use of methods with low computational cost but near optimal outputs in the second phase. According to the fact that the length of the final output has a direct effect on the cost of performing the task, it is desirable to incorporate methods with low complexity that can guarantee a bound for the difference between length of the optimal path and

the output. Current approaches for solving subpath planning problem are all meta-heuristic based. These methods do not provide such a bound. And plus, they are usually very time consuming. They may find promising results for some instances of problems, but there is no guarantee that they always exhibit such a good behaviour. In this paper, in order to avoid the issues of meta-heuristics methods, we present an approximation algorithm, which provides an appropriate bound for the optimality of its solution. To gauge the performance of proposed methods, we conducted a set of experiments the results of which show that our proposed method finds shorter paths in less time in comparison with a state-of-the-art method.

1 Introduction

Path planning is a challenging problem in artificial intelligence and robotics [30] with applications also in other areas such as computer animation and computer games [14], therapeutic [5] protein folding [35], manufacturing process [34] and computational biology [36]. Due to various applications, different types of this problem have been proposed including the *subpath planning problem* (SPP). SPP has widespread applications such as navigation of robots and vehicles as well as automated manufacturing process [16, 37]. The goal of SPP is to find the shortest tour, which travels all given subpaths. SPP is an NP-hard problem [13, 19]. As an example, Fig. 1.a shows a workspace and Fig. 1.b represents its corresponding optimal result. Also, Fig 1.a represents the corresponding graph constructed based on the workspace.

There have been proposed a few approaches for solving SPP, all of which are meta-heuristic based. Recently, Ying *et al.* [37] and Györfi *et al.* [16] proposed some algorithms based on *Genetic Algorithm* (GA) [15] for solving SPP in polishing robots and electronic printing, respectively.

Like other meta-heuristic methods [4], GA cannot guarantee any bound on its final result. It may produce some promising results on some given instances, while it has a tendency to converge to local optima for some other instances. In addition, GA needs considerable amount of time in order to return a result. This problem becomes more severe as the number of subpaths grows.

The current paper aims at overcoming the problems of meta-heuristic methods in solving SPP by proposing an approximation algorithm [3] with a fixed ratio bound and efficient polynomial complexity. Our method includes three stages described as follows.

The first stage is transforming SPP to *Travelling Salesman Problem* (TSP) [18] with an $O(n^2)$ time complexity algorithm. TSP is a well-known combinatorial optimization problem. Since TSP is an NP-hard problem [19], proposing a precise algorithm for solving TSP does not make sense. Thus, several attempts have been done to propose approximation algorithms for solving this problem. In the recent decades,

various approximation methods have been proposed for solving this problem.

Once an SPP instance is transformed to a suitable TSP one, it may seem easy to apply the existing fixed-ratio bound approximation algorithms for TSP for solving SPP. However, this is not the case and there are some crucial challenging issues in this way. Christofides in [6] argues that there is no polynomial approximation algorithm with a fixed ratio bound for general TSP. Literally, we cannot propose any fixed ratio bound approximation algorithm on a general graph, in which the triangle inequality does not hold in all triangles [6] (this observation is also due to Sahni and Gonzale [33]). However, there are some fixed-ratio bound approximation algorithms such as *Christofides'* algorithm [6], for solving TSP over constrained graphs, which satisfy triangle inequality. Here is the point where we face a crucial problem: the output graph of transforming of SPP to TSP for a given instance definitely violates the triangle inequality condition, as shown in Section 3. Thus, it is not feasible to apply existing fixed-ratio bound TSP approximation algorithms for solving SPP. We address this shortcoming in a two next stages.

In the second stage, we propose an algorithm, called *Imperfectly Establish the Triangle Inequality* (IETI), which establishes the triangle inequality in a main subset of violating triangles¹. The output of the first stage, i.e., transforming SPP to TSP, is considered as the input graph of the IETI algorithm. The algorithm outputs a new graph by changing the edges' weight of the input such that the triangle inequality condition holds on all triangles except for some special triangles (those that one and only one of their edges does have the infinity weight). Let G' be the result of transforming an SPP instance graph G and G'' be the output graph of the IETI algorithm for G' . We formally show that solving SPP on G would be equivalent to solving TSP on G'' . The IETI algorithm is in $O(n^2)$ complexity class and should be seen as a fundamental step for introducing and applying a fixed-ratio bound approximation algorithm for solving SPP. This is because, as discussed already, the main requirement of applying such an algorithm for solving SPP is holding triangle inequality in the given graph.

Nonetheless, some special triangles still violate the triangle inequality in the output graph of the IETI algorithm. To tackle this problem, in the third stage, we propose an approximation algorithm with the fixed-ratio bound of 2 and $O(n^3)$ complexity, called *Christofides for SPP* (CSPP). Indeed, the CSPP algorithm is a modified version of the Christofides' algorithm [6] to make it able to work for all outputs of the IETI algorithm. The Christofides' algorithm is a polynomial approximation algorithm with fixed-ratio bound of 1.5² for solving TSP instances in which edge

¹As it may be clear, a violating triangle is a triangle which violates the triangle inequality condition.

²The fixed ratio bound of 1.5 is the minimum among the existing methods proposed for solving TSP.

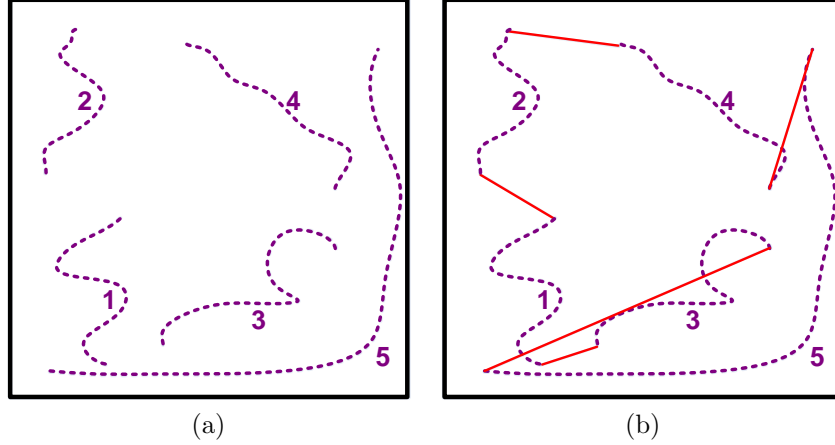


Figure 1: (a) The workspace related to a subpath planning problem where dashed curves represent the subpaths. (b) The optimal solution of the corresponding SPP.

weights are metric³. Thus, any input graphs of this algorithm must satisfy the triangle inequality condition. Therefore, it is not feasible to apply Christofides to our problem. In other words, the CSPP algorithm aims at solving TSP for given graphs in which the triangle inequality holds in all triangles except for those that one of their edges has the infinity weight.

In addition to complexity analysis and proving the ratio bound of CSPP, it is empirically compared with the method proposed by Györfi *et al.* [16] over various workspaces with different number of subpaths. The results illustrate that CSPP is more efficient than the state-of-the-art method in terms of both result and running time.

The rest of the paper is organized as follows. Section 2 discusses the related work. Section 3, presents transformation of SPP to TSP. We discuss the IETI algorithm in Section 4. In Section 5, we propose the CSPP algorithm and describe its characteristics and the related theorems. The experimental comparison of CSPP and the method proposed by Györfi *et al.* [16] is presented in Section 6. Finally, the conclusions and future work are discussed in Section 7.

2 Related Work

There exist some graph problems, which are relevant to SPP. This set of problems includes *Travelling Salesman Problem with Neighbours* (TSPN) and those that are in the context of *Arc Routing Problems* (ARP) [9, 10]. Below, we discuss their similar-

³like other fixed-ratio bound algorithms for solving the TSP

ities and differences with SPP.

Travelling Salesman Problem with Neighbours: *Travelling Salesman Problem with Neighbours* (TSPN) is introduced by Arking and Hassin [2]. It is a generalization of TSP in which the constraint is to visit the neighbourhood of each node instead of the node itself. In TSPN, each node is represented as a polygon instead of a single point and an optimal solution is the shortest path such that it intersects all polygons. Since TSPN is a generalization of TSP, it is also NP-hard [26, 28]. Besides, Safra and Schwartz [32] showed that it is NP-hard to approximate within any constant bound. For the general case of connected polygons, Mata and Mitchell [24] proposed an $O(\log n)$ approximation bound with $O(N^5)$ time complexity based on "guillotine rectangular subdivisions", where N is the total number of vertices of the polygons. If all the polygons have the same diameter, then an $O(1)$ algorithm also exists [8]. Even if we represent each subpath with a polygon of two vertices, then SPP is different from TSPN. This is because an SPP solution requires to traverse all the subpaths, while a solution for TSPN can only have intersections with each subpaths.

Rural Postman Problem: Consider a graph $G(V, E)$, where V is the set of vertices and E is the set of edges. In the *Chinese Postman Problem* (CPP), we are interested in finding the shortest closed path such that it travels all the edges. An optimal solution is an Eulerian tour, if exists any. Thus, whenever the degree of each node is even, CPP can be reduced to finding an Eulerian tour. Note that it is well-known that an Eulerian tour always exists in such a graph. If G is either purely directed or purely undirected, CPP has a polynomial time solution. Otherwise (the given graph is neither purely directed nor undirected), the problem would be NP-hard [10].

Rural Postman Problem (RPP) is a variant of CPP. A CPP problem is called RPP, if a subset of edges must be covered instead of covering all the edges. RPP was first introduced by Orloff [27]. The undirected, directed and mixed versions of RPP are all proven to be NP-hard [11, 22]. Frederickson [11] proposed a polynomial time solution for RPP with the worst case ratio bound of 1.5 for given graphs, which satisfy the triangle inequality condition. This solution is known as the best one for RPP.

There are many similarities between RPP and SPP. Indeed, RPP is a generalization of SPP. An SPP instance is an RPP instance in which the subpaths are the edges that must be covered. As Fig. 2 shows, there are additional constraints in SPP. The number of vertices is twice the number of edges that must be covered (subpaths). Thus, the must-be-covered edges in an SPP instance cannot share a common vertex. Besides, the graph is an undirected complete one. Although SPP and RPP have many similarities, fixed-ratio bounds algorithms for RPP cannot be applied for SPP. This is because given graphs for SPP do not satisfy the triangle inequality condition.

Stacker Crane Problem: The *Stacker Crane Problem* (SCP) [12] is another relevant problem in the context of routing. SCP is defined on a graph consisting of directed and undirected edges. The problem is to find the shortest circuit, which covers all the directed edges (which can be the deliveries that to be made by a vehicle). SCP is also an NP-hard problem, since it is a generalization of TSP. Coja-Oghlan et al. [7] proposed an approximation approach for a special case of SCP. In this solution, given graphs must be trees. Even in such a restricted case, the problem is NP-hard. Fredrickson et al. [12] proposed a polynomial algorithm for this problem with the ratio bound of 1.8 in the worst case. The Fredrickson’s solution for SCP is known to be the best approximation algorithm [38]. The difference between SCP and SPP is that, in SPP, subpaths, which must be covered, are indirected edges.

3 Transformation of SPP to TSP

In this section, we show how to transform SPP to TSP. Feasible solutions for an SPP instance are the tours that travel all the subpaths. A tour with the minimum length is a desired solution. Each feasible solution is a sequence of connected subpaths. In a fixed sequence with n subpaths, the i^{th} ($i < n$) subpath can be connected to the $(i + 1)^{\text{th}}$ subpath (the n^{th} subpath is connected to the first one) in two different ways (either to head or tail). Now, consider an SPP instance with n subpaths. Obviously, the number of possible sequences of these n subpaths is $n!$. Thus, due to two different ways of connections between two consecutive subpaths, the total number of feasible solutions would be $n!2^n$.

TSP is one of the classical NP-hard problems of combinatorial optimization. In the recent decades, various approximation [21] and combinatorial optimization methods [18] have been proposed for solving this problem. Thus, transformation of SPP to TSP facilitates applying such methods for solving SPP. The rest of the section is organized as follows. The subsection 3.1 discusses the transformation procedure of SPP to TSP and in the subsection 3.2, we discuss the complexity analysis of the procedure on its corresponding pseudo code.

3.1 TSP model of SPP: Transformation Procedure

Consider an SPP instance with n subpaths indexed with the set $I = \{1, \dots, n\}$. The procedure includes two stages. In the first stage, a complete graph G is built, according to the following stages.

Stage 1:

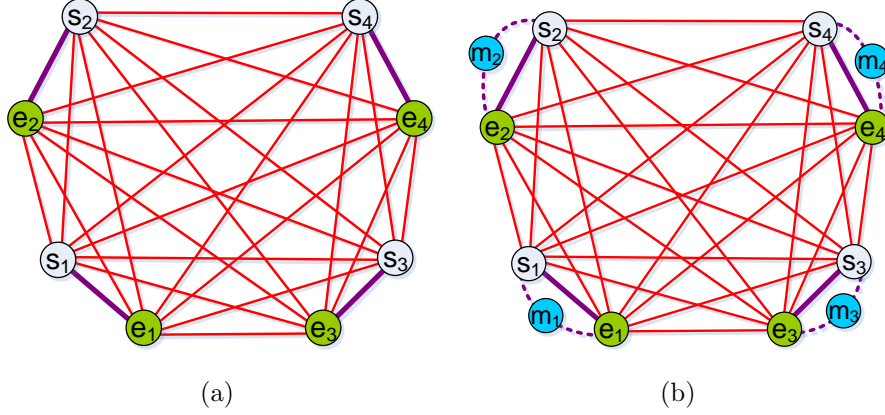


Figure 2: (a) graph G represents the workspace. In this graph, thick edges indicate subpaths and thin edges indicate the connections between the subpaths. The weight of connecting edges are equal to their corresponding Euclidean distances between subpaths in the workspace. (b) The graph G' is built by adding middle nodes to G . Infinity weight edges are not shown here.

1. For each subpath, say i^{th} ($i \in I$), consider two nodes s_i and e_i corresponding to its starting and end points, respectively.
2. For each $i \in I$, consider an edge between s_i and e_i with the weight equal to the length of i^{th} subpath in the workspace. Let us call this edge the i^{th} *subpath edge*.
3. For each pair of two distinct subpaths i and j ($i \neq j$), we add edges $s_i e_j$, $e_i s_j$, $s_i s_j$ and $e_i e_j$ to the graph. We also consider the weights of these newly added edges equal to the corresponding Euclidean distances in the workspace. Let us call these edges the *connecting edges*.

Fig. 2.a depicts the graph G generated according to Stage 1 for the given workspace in Fig. 1.a.

The TSP tour of G (the graph generated in the above procedure) is not necessarily equivalent to the solution of the given SPP instance. This is because the solution of SPP is a tour traveling all the subpaths, while the TSP tour of G may not cover all the subpath edges (such as $s_i e_i$). To make sure that the TSP tour travels all the subpaths, a complete graph G' is generated based on G as follows:

Stage 2:

1. For each subpath, say i^{th} subpath, a node m_i is added to the graph, called the *middle node* of i^{th} subpath.

2. For each subpath, say i^{th} , two edges $s_i m_i$ and $m_i e_i$ are added to the graph, the weights of which are equal to the half of the subpath length. These edges are called i^{th} double subpath edges.
3. For each middle node m_i , the edge $m_i v$, where $v \notin \{s_i, e_i\}$, is added to the graph with the infinity weight.

Fig. 2.b depicts the graph G' generated according to Stage 2 for the given graph G in Fig. 2.a.

Theorem 1 *The result of solving SPP on a given instance is equivalent to finding the TSP tour in G' generated according to the above procedures (Stage 1 + Stage 2) on the instance.* ■

Proof: According to Fig. 2.b, there is a finite Hamiltonian tour in G' . $[s_1-m_1-e_1-s_2-m_2-e_2-s_3 \dots s_n-m_n-e_n]$ is a sample of finite hamiltonian tours in G' with n subpaths.

The TSP tour over G' is a Hamiltonian tour with the minimum weight. Therefore, The TSP tour over G' is finite. The TSP tour of G' must visits all the middle nodes, since, for each i , it contains two edges crossing the node m_i . There are only two finite edges s_i-m_i and m_i-e_i connecting to m_i . Hence, the TSP tour over G' must contain i^{th} double subpath edges for each i . Since $s_i m_i$ and $m_i e_i$ together are equivalent to i^{th} subpath in the workspace, the TSP tour of G' is a minimum tour, which travels all subpaths. Hence, solving SPP is equivalent to finding the TSP tour in G' . □

Throughout the rest of the paper, we use the notation G' to denote the graph generated in the above procedure for a given SPP instance.

3.2 Pseudo Code and Complexity Analysis of Transforming

Algorithm 1 presents a pseudo code for the SPP to TSP transformation procedure. The algorithm takes a workspace as input and returns a graph G' as output. It includes the following two phases:

- 1) Generating a graph G , according to Stage 1 (Line 1)
- 2) Generating a graph G' by adding middle nodes to G , according to Stage 2 (Lines 2 to 9)

Time complexity of generating the graph G is in $O(n^2)$, where n is the number of subpaths. Lines 2 to 9 add middle nodes to the graph within a loop of n iterations. In each iteration, there is another loop (lines 6-8), which requires $O(n)$ running time. Therefore, adding middle nodes requires $O(n^2)$ running time. Thus, the total complexity of the algorithm is $O(n^2)$.

Algorithm 1 : SPP to TSP

```
1: Construct  $G$  with the adjacency matrix  $w$ 
2: for  $i = 1$  to  $n$  do
3:   add middle node  $m_i$ 
4:    $w(m_i, s_i) \leftarrow \frac{w(e_i, s_i)}{2}$ 
5:    $w(m_i, e_i) \leftarrow \frac{w(e_i, s_i)}{2}$ 
6:   for each node  $d \in G$  where  $d \notin \{s_i, e_i\}$  do
7:      $w(m_i, d) \leftarrow \infty$ 
8:   end for
9: end for
```

4 Imperfectly Establish the Triangle Inequality

As discussed already, using any existing approximation method for TSP requires triangle inequality to hold over given graphs. Two kinds of triangles in G' (Fig. 2.b) may violate triangle inequality:

V-1) Triangles with a subpath edge as one of their edges, i.e, triangles in the form of $\triangle s_i e_i v$ ⁴, where $v \neq m_i$.

V-2) Triangles with one and only one infinity edge. In such a triangle, one of its edges is either $s_i m_j$ or $e_i m_j$ ($i \neq j$).

Other triangles in G' that are not in one of the above kinds do not violate the triangle inequality condition. Such triangles can be grouped into the following kinds:

- 1) Triangles that have more than one infinity edge.
- 2) Triangles in which all of the edges' weights are equal to their corresponding Euclidean distances.
- 3) Triangles that are of the form $\triangle s_i m_i e_i$.

As mentioned in the introduction, we tackle the triangle inequality violation in G' in two stages. The first stage is proposing an algorithm, called Imperfectly Establish the Triangle Inequality (IETI). We discuss how the algorithm works in the subsection 4.1. The subsection 4.2 discusses the complexity analysis of the procedure on its corresponding pseudo code.

4.1 IETI: The Procedure

The IETI algorithm is given the output of the transformation procedure (G') and deals with the first category of violating triangles, i.e., **V-1**. Indeed, this algorithm makes some modifications to the edges' weight of G' to make a graph, denoted by G'' , such that TSP tours in G' and G'' are the same (see Theorem 3) and there is no violating triangles of kind **V-1** in G'' (see Theorem 2).

⁴ $\triangle abc$ denotes the triangle with a, b , and c as its vertices.

The IETI algorithm is an iterative method. Indeed, it iterates over all subpaths, for each of which it updates the weight of edges in a same way. Indeed, each iteration corresponds to a subpath. Below, we describe how it works.

For each iteration, say i^{th} (corresponding to the i^{th} subpath), we define a variable called i^{th} degree of violation, denoted by dv_i . We apply this variable to formally recognize what triangles in **V-1** violate the triangle inequality condition. We also use it to resolve such violations. The equation (1) shows how to compute dv_i .

Remark 1 For a given graph H , the notations $V(H)$ and $w(a, b)$ denote the set of vertices and the weight of the edge ab , respectively.

$$\begin{aligned} \forall d \in V(G') - \{s_i, e_i\} \\ dv_i = 0.5(\min_{\forall d} (w(s_i, d) + w(e_i, d)) - w(s_i, e_i)) \end{aligned} \quad (1)$$

$dv_i < 0$ implies that at least one of the triangles in which one of their edges is $e_i s_i$ violates the triangle inequality condition. Otherwise, i.e., $dv_i \geq 0$, none of such triangles violates the condition. In the former case, the weight of edges are updated by the following equations:

$$\begin{aligned} w(s_i, e_i) &\leftarrow w(s_i, e_i) - |dv_i| \\ w(s_i, m_i) &\leftarrow w(s_i, m_i) - \left|\frac{dv_i}{2}\right| \\ w(e_i, m_i) &\leftarrow w(e_i, m_i) - \left|\frac{dv_i}{2}\right| \end{aligned} \quad (2)$$

$$\begin{aligned} \forall q \in V(G') - \{s_i, e_i, m_i\} \\ w(q, e_i) &\leftarrow w(q, e_i) + \frac{|dv_i|}{2} \\ w(q, s_i) &\leftarrow w(q, s_i) + \frac{|dv_i|}{2} \end{aligned} \quad (3)$$

Note that, in equation 3, the added weights to the edges, which are connected to the subpaths, are equal and symmetric, i.e., the weights of edges connected to s_i and e_i are increased equally. This property makes the TSP tour over G' to be equivalent to one over G'' (proven in Theorem 3).

These changes make the triangles containing the i^{th} subpath to satisfy the triangle inequality. Note that it does not make other triangles, which already satisfy the triangle inequality, to violate the condition. This claim is proven in Theorem 2.

Theorem 2 shows that IETI establishes the triangle inequality in all **V-1** triangles in G'' such that other triangles except for those in **V-2** still satisfy the triangle inequality. Throughout the rest of the paper, we use the notation G'' to denote the graph generated in the above procedure (IETI) for G' .

Theorem 2 *After the execution of IETI, all the triangles in G'' satisfy the triangle inequality except for those in **V-2**. ■*

Proof: Let n be the number of subpaths of the original workspace. Consider the i^{th} step of IETI. If $dv_i < 0$ (computed in the equation 1), then, according to the equations (2) and (3), the weights of the corresponding edges change. Let G''_i denote the result graph after the execution of the i^{th} step of IETI on G' . It is also natural to consider G''_0 and G''_n equal to G' and G'' , respectively.

Now, we are going to prove the following statement:

Statement: *A violating triangles in G''_i is either in:*

*g-1) **V-2** or*

*g-2) **V-1** such that it is a triangle $\triangle s_j e_j v$ with $j > i$.*

We use an inductive reasoning to prove the above statement, as follows.

(*base case*): It follows obviously that the above statement holds in G''_0 (which is equal to G').

(*hypothesis*): Assume that, for some t with $1 \leq t < i$, the statement holds for G''_0, \dots, G''_{t-1} , now it suffices to show that the statement also holds for G''_t . This is shown in the inductive step.

(*inductive step*): There exists the two following possible cases for G''_{t-1} . We show in each case the G''_t satisfies the statement.

1. “In G''_{t-1} , the triangles with the edge $s_t e_t$ do not violate the triangle inequality condition.”

Thus, for any $j \leq t$, the triangles in G''_{t-1} with an edge $s_j e_j$ do not violate the triangle inequality. In this case, during the t^{th} step, no modification will be made to the graph G''_{t-1} and G''_t would be equal to G''_{t-1} . Thus, for any $j \leq t$, the triangles in G''_t with an edge $s_j e_j$ do not violate the triangle inequality. Hence the statement holds for G''_t .

2. “There are some triangles in G''_{t-1} with an edge $s_t e_t$, which violates the triangle inequality.”

Consider a violating triangle in G''_{t-1} with an edge $s_t e_t$. Let us see what would happen in the t^{th} step. According to the equation (2), the weight of the edge $s_t e_t$ must decrease by $|dv_t|$. Moreover, for any q , the weights of the edges $s_t q$ and $q e_t$ increase by $\frac{|dv_t|}{2}$ (equation 3).

Each triangle in G''_t , say $\triangle abc$, can fall into one of the seven following categories. The validity of triangle inequality in all categories will be investigated. In other words, the validity of the inequalities $w_t(a, c) + w_t(b, c) \geq w_t(a, b)$ and

$w_t(a, b) + w_t(b, c) \geq w_t(a, c)$, where w_t is the adjacency matrix of G_t'' and $w_t(a, b)$ denotes the weight of the edge ab in G_t'' ⁵.

(a) “The triangle edge ab is equal to $s_t e_t$.”

In this case, the weight of the edge ab decreases by $|dv_t|$ and the weights of the edges ac and bc increases by $\frac{|dv_t|}{2}$. The equations (5) and (9) show that the triangles in this category satisfy the triangle inequality after modifications:

$$\begin{aligned}
& w_t(a, c) + w_t(b, c) \\
&= w_t(s_t, c) + w_t(e_t, c) \\
&= w_{t-1}(s_t, c) + \frac{|dv_t|}{2} + w_{t-1}(e_t, c) + \frac{|dv_t|}{2} \\
&= w_{t-1}(s_t, c) + w_{t-1}(e_t, c) + 2|dv_t| - |dv_t|
\end{aligned} \tag{4}$$

By replacing $2|dv_t|$ with $w_{t-1}(s_t, e_t) - \min_{\forall d}[w_{t-1}(s_t, d) + w_{t-1}(e_t, d)]$, we would have:

$$\begin{aligned}
& w_t(a, c) + w_t(b, c) \\
&= w_{t-1}(s_t, c) + w_{t-1}(e_t, c) + w_{t-1}(s_t, e_t) \\
&\quad - \min_{\forall d}[w_{t-1}(s_t, d) + w_{t-1}(e_t, d)] - |dv_t| \\
&\geq w_{t-1}(s_t, e_t) - |dv_t| = w_t(s_t, e_t) = w_t(a, b)
\end{aligned} \tag{5}$$

$$\begin{aligned}
& w_t(a, b) + w_t(b, c) \\
&= w_t(s_t, e_t) + w_t(e_t, c) \\
&= w_{t-1}(s_t, e_t) - |dv_t| + w_{t-1}(e_t, c) + \frac{|dv_t|}{2}
\end{aligned} \tag{6}$$

By replacing $|dv_t|$ with $\frac{1}{2}(w_{t-1}(s_t, e_t) - \min_{\forall d}[w_{t-1}(s_t, d) + w_{t-1}(e_t, d)])$, we

⁵Validity investigation of $w_t(a, b) + w_t(a, c) \geq w_t(b, c)$ is similar to of $w_t(a, b) + w_t(b, c) \geq w_t(a, c)$. So, there is no need to investigate the validity of the former inequality in the proof.

would have:

$$\begin{aligned}
w_t(a, b) + w_t(b, c) &= w_{t-1}(s_t, e_t) + w_{t-1}(e_t, c) \\
&\quad - \frac{w_{t-1}(s_t, e_t) - \min_{\forall d} [s_{t-1}(s_t, d) + s_{t-1}(e_t, d)]}{2} \\
&\quad + \frac{|dv_t|}{2} \\
&= \frac{w_{t-1}(s_t, e_t)}{2} \\
&\quad + \frac{\min_{\forall d} [w_{t-1}(s_t, d) + w_{t-1}(e_t, d)]}{2} \\
&\quad + w_{t-1}(e_t, c) + \frac{|dv_t|}{2}
\end{aligned} \tag{7}$$

The weight of the edge $s_t e_t$ has not changed before the t^{th} step. Thus, $w_{t-1}(s_t, e_t) = w_0(s_t, e_t)$. In other words, it is equal to the length of t^{th} subpath. Hence, $w_{t-1}(s_t, e_t) > ed(s_t, e_t)$, where $ed(s_t, e_t)$ is the Euclidean distance between s_t and e_t in the workspace.

Also, any node such as d satisfies the inequality $w_{t-1}(s_t, d) + w_{t-1}(e_t, d) \geq ed(s_t, d) + ed(e_t, d) > ed(s_t, e_t)$.⁶ As a consequence, the inequality $\min_{\forall d} [w_{t-1}(s_t, d) + w_{t-1}(e_t, d)] > ed(s_t, e_t)$ holds, which results in:

$$\begin{aligned}
w_t(a, b) + w_t(b, c) \\
\geq ed(s_t, e_t) + w_{t-1}(e_t, c) + \frac{|dv_t|}{2}
\end{aligned} \tag{8}$$

If the weights of subpaths, which pass through the node c change before the t^{th} step, then the values of $w_{t-1}(s_t, c)$ and $w_{t-1}(e_t, c)$ would be equal to $ed(s_t, c) + \frac{\alpha}{2}$ and $ed(c, e_t) + \frac{\alpha}{2}$, respectively, where α is equal to the parameter dv (degree of violation) of the subpath, which passes through the node c .

Otherwise (if the value of such subpaths are not updated), the values of $w_{t-1}(s_t, c)$ and $w_{t-1}(e_t, c)$ would be $ed(s_t, c)$ and $ed(c, e_t)$, respectively. Hence, the inequality $ed(s_t, e_t) + w_{t-1}(c, e_t) > w_{t-1}(s_t, c)$ turns to $ed(s_t, e_t) + ed(c, e_t) > ed(s_t, c)$ that always holds. Consequently, the following inequal-

⁶The weights of edges $s_t d$ and $e_t d$ either have not changed before t^{th} step or are modified by equation (3). Therefore, $w_{t-1}(s_t, d) \geq w_0(s_t, d) \geq ed(s_t, d)$ and $w_{t-1}(e_t, d) \geq w_0(e_t, d) \geq ed(e_t, d)$.

ity holds:

$$\begin{aligned}
& w_t(a, b) + w_t(b, c) \\
& \geq w_{t-1}(s_t, c) + \frac{|dv_t|}{2} \\
& \geq w_t(s_t, c) \\
& \geq w_t(a, c)
\end{aligned} \tag{9}$$

(b) “The edge ab is $s_j e_j$ (i.e., $a = s_j$ and $b = e_j$), where $j < t$ ”.

In this case, $\triangle abc$ in G''_{t-1} satisfies the triangle inequality. Two following scenarios are possible during the step t^{th} :

- 1) The edges' weight of $\triangle abc$ do not change in t^{th} step. In this case, obviously, $\triangle abc$ still satisfies the triangle inequality in t^{th} step.
- 2) The weights of the edges ac and bc increase by $\frac{|dv_t|}{2}$ (Note that this scenario happens when c is either s_t or e_t). In this case, the following equations show that $\triangle abc$ still satisfies the triangle inequality:

$$\begin{aligned}
& w_t(a, c) + w_t(b, c) \\
& = w_t(s_j, c) + w_t(e_j, c) \\
& = w_{t-1}(s_j, c) + \frac{|dv_t|}{2} \\
& \quad + w_{t-1}(e_j, c) + \frac{|dv_t|}{2} \\
& \geq w_{t-1}(s_j, c) + w_{t-1}(e_j, c) \\
& \geq w_{t-1}(s_j, e_j) = w_t(s_j, e_j) = w_t(a, b)
\end{aligned} \tag{10}$$

and

$$\begin{aligned}
w_t(a, b) + w_t(b, c) & = w_t(s_j, e_j) + w_t(e_j, c) \\
& = w_{t-1}(s_j, e_j) + w_{t-1}(e_j, c) \\
& \quad + \frac{|dv_t|}{2} \\
& \geq w_{t-1}(s_j, c) + \frac{|dv_t|}{2} \\
& = w_t(s_j, c) \\
& = w_t(a, c)
\end{aligned} \tag{11}$$

(c) “The edge ab is $s_j e_j$, where $j > t$ ”.

In this case, the triangle $\triangle abc$ may violate the triangle inequality in the graph G''_{t-1} . Hence, it may violate the triangle inequality in G''_t too.

- (d) “ $\triangle abc$ does not have any edge in the form of $s_j e_j$ for some j (i.e., it does not have any subpath edge), and the weights of the edges ac and bc change during the t^{th} step”.⁷

In this case, the triangle $\triangle abc$ satisfies the triangle inequality condition in the graph G''_{t-1} . During the t^{th} step, the weights of the edges ac and bc increase by $\frac{dv_t}{2}$. The following equations show that this triangle still satisfies the triangle inequality in G''_t :

$$\begin{aligned} w_t(a, c) + w_t(b, c) &= w_{t-1}(a, c) + \frac{dv_t}{2} \\ &\quad + w_{t-1}(b, c) + \frac{|dv_t|}{2} \\ &\geq w_{t-1}(a, c) + w_{t-1}(b, c) \\ &\geq w_{t-1}(a, b) = w_t(a, b) \end{aligned} \tag{12}$$

$$\begin{aligned} w_t(a, b) + w_t(b, c) &= w_{t-1}(a, b) + w_{t-1}(b, c) \\ &\quad + \frac{|dv_t|}{2} \geq w_{t-1}(a, c) \\ &\quad + \frac{|dv_t|}{2} = w_t(a, c) \end{aligned} \tag{13}$$

- (e) “ $\triangle abc$ does not have any edge in the form of $s_j e_j$ for some j (i.e., it does not have any subpath edge) and the weight of no edge of the triangle change during the t^{th} step”.

Since $\triangle abc$ satisfies the triangle inequality condition in G''_{t-1} , it also satisfies the inequality in G''_t .

- (f) “ $\triangle abc$ has one edge with the infinity weight”.

The triangle violates the triangle inequality condition in both G''_{t-1} and G''_t .

- (g) “ $\triangle abc$ has two or three edges with the infinity weight”.

It follows obviously that $\triangle abc$ satisfies the triangle inequality condition in both G''_{t-1} and G''_t .

Thus, after the execution of the t^{th} step, only the triangles with one infinity weight edge (category f) or with edges in the form of $s_j e_j$ for some j greater than t (category c) may violate the triangle inequality in G''_t . Therefore, the statement holds in G''_t .

The statement was proven, which means that it holds in $G''_n = G''$.

⁷Note that during each step of IETI for each triangle either no weight is updated (category e) or two weights are updated (category d).

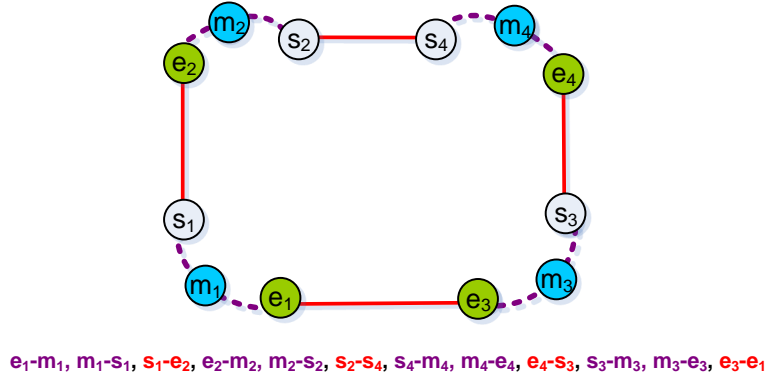


Figure 3: Example of Hamiltonian tour over G' and G'' . The pair of edges are shown by dashed arrows and other arrows represent the connecting edges between the pair of edges.

As a result, the triangle inequality condition is satisfied by every triangle in the graph G'' (the output graph of IETI) except for those with an infinity edge. These triangles are either in the form of $\triangle s_i m_i d$ or $\triangle e_i m_i d$, where the edge $m_i d$ is an infinite edge. The theorem is proven. \square

The following theorem shows that the TSP tours in a given graph of IETI, i.e., G' (the result of SPP to TSP transformation for a given SPP workspace) and the output graph of IETI, i.e., G'' , are the same. This implies that the TSP tour in G'' is equivalent to the SPP solution of the given workspace.

Theorem 3 *The TSP tours in G' and G'' are the same.* ■

Proof: To prove this theorem, we show that the TSP tours of G' and G'' are the same in terms of length and sequence of nodes. In other words, we show that the length of each finite Hamiltonian tour of G' is equal to the length of its corresponding finite Hamiltonian tour of G'' (A Hamilton tour over G' and another one over G'' corresponds to each other, if they have the same sequence of nodes).

As already stated, any finite Hamiltonian tours in both G' and G'' include all edges is the form of $s_i m_i$ and $m_i e_i$ (for any possible index i). Note that in such a tour, for any i , $m_i e_i$ and $s_i m_i$ happen consecutively and are connected through m_i (as a sequence in the form of either $s_i-m_i-e_i$ or $e_i-m_i-s_i$). Let us call these two consecutive edges "pair of edges of i^{th} subpath". As a result, each finite Hamiltonian tour in either G' or G'' is a sequence of pairs of edges connected via some edges. Fig. 3 presents an example of Hamiltonian tour over G' and G'' .

Note that the graph G' differs from the graph G'' in their weights of the edges. Let H be a finite Hamiltonian tour over G' and H' be its corresponding finite Hamiltonian tour over G'' (suppose that the number of subpaths is n). Without loss of generality,

we let the pairs of edges occur in H (also in H') according the usual order of their indices, i.e., for any index i less than n , the $(i + 1)^{\text{th}}$ pair happens exactly after the i^{th} pair of edges. Suppose that the corresponding node p in G' is p' in G'' .

$$H : e_1 m_1, m_1 s_1, s_1 e_2, \dots, s_{i-1} e_i, e_i m_i, m_i s_i, s_i e_{i+1}, \dots, \\ e_{n-1} s_n, s_n m_n, m_n e_n$$

and

$$H' : e'_1 m'_1, m'_1 s'_1, s'_1 e'_2, \dots, s'_{i-1} e'_i, e'_i m'_i, m'_i s'_i, s'_i e'_{i+1}, \dots, \\ e'_{n-1} s'_n, s'_n m'_n, m'_n e'_n$$

During the i^{th} of the IETI procedure over G' , only the weights of the edges in H , which are in the form of $e_i m_i$, $m_i s_i$, $s_i e_i$, $e_i s_{i-1}$, and $s_i e_{i+1}$, may change and others are left without any changes. Thus, the following equations hold:

$$w(e_i m_i) + w(m_i s_i) = w(e'_i, m'_i) + w(m'_i, s'_i) + |dv_i| \quad (14)$$

$$w(s_{i-1}, e_i) = w(s'_{i-1}, e'_i) - \frac{|dv_i|}{2} \\ w(s_i, e_{i+1}) = w(s_i, e_{i+1}) - \frac{|dv_i|}{2} \quad (15)$$

The above equations show that changes made in weights in each step of IETI do not make the length of the corresponding tours H and H' unequal. Therefore, the length of each finite Hamilton tour in G' is equal to its corresponding tour in G'' . \square

4.2 Pseudo Code and Complexity Analysis of IETI

Algorithm 2 presents the pseudo code of the IETI algorithm. This algorithm takes a graph G' (the output of the SPP to TSP transformation) and returns a graph G'' .

The loop of n iterations in lines 1 to 13 updates the weights of the subpath edges. There exist two loops within these lines each of which iterates $2n$ times (lines 2 and 8 to 11). Thus, the time complexity of the algorithm is in the $O(n^2)$ class.

5 A 2-approximation Algorithm

The Christofides' algorithm [6] is one the most efficient approximation algorithms for solving TSP, which works for given graphs satisfying the triangle inequality condition. This algorithm has time complexity of $O(n^3)$ and the ratio bound of 1.5. Due to its complexity and ratio bound, the Christofides' algorithm is a popular approximation method for TSP.

Algorithm 2 : TSP Transformation and IETI

```
1: for each subpath edge  $s_i e_i$  do
2:   Find node  $d \in G$  that minimizes  $R = w(s_i, d) + w(d, e_i)$ 
3:    $dv \leftarrow \frac{R - w(s_i, e_i)}{2}$ 
4:   if  $dv < 0$  then
5:      $w(s_i, e_i) \leftarrow w(s_i, e_i) - |dv|$ 
6:      $w(s_i, m_i) \leftarrow \frac{w(s_i, e_i)}{2}$ 
7:      $w(e_i, m_i) \leftarrow \frac{w(s_i, e_i)}{2}$ 
8:     for each node  $q \in G$  where  $q \notin \{s_i, e_i, m_i\}$  do
9:        $w(s_i, q) \leftarrow w(s_i, q) - \frac{|dv|}{2}$ 
10:       $w(e_i, q) \leftarrow w(e_i, q) - \frac{|dv|}{2}$ 
11:     end for
12:   end if
13: end for
```

According to Theorem 3, the TSP tour in G'' is equivalent to the solution of SPP. However, G'' violates the triangle inequality condition. Therefore, it is not feasible to use Christofides' algorithm (or any other existing fixed-ratio bound approximation algorithms) to find the TSP tour in G'' . Nonetheless, the triangle inequality is violated by some special triangles in G'' . By using this special feature, an approximation algorithm for finding the TSP tour over G'' , called *Christofides for SPP* (CSPP), is proposed with $O(n^3)$ time complexity and ratio bound of 2. CSPP can be seen as a modified version of Christofides' algorithm.

The CSPP algorithm contains one additional step in comparison with the Christofides' algorithm. Moreover, one step of Christofides is modified in CSPP.

The plan of this section is as follows. In 5.1, we discuss the CSPP algorithm. In 5.2, we show that the ratio bound of the CSPP algorithm is 2. We finally discuss the time complexity of the algorithm in 5.3.

5.1 CSPP: the procedure

CSPP takes G'' (the output of the IETI algorithm) as input and returns a Hamiltonian tour as output. This algorithm consists of five steps as follows:

Step 1: Finding the Minimum Spanning Tree

This step finds the minimum spanning tree (MST) over G'' (one of the nodes is arbitrarily chosen as the root of the tree). This step is the same as the first step of the Christofides' algorithm [6].

Step 2: Modify MST by Adding Subpath Edges

The MST (the result of the first step) does not include an edge with infinity weight. As a consequence, a middle node of the graph G'' , say m_i for some i , can be either

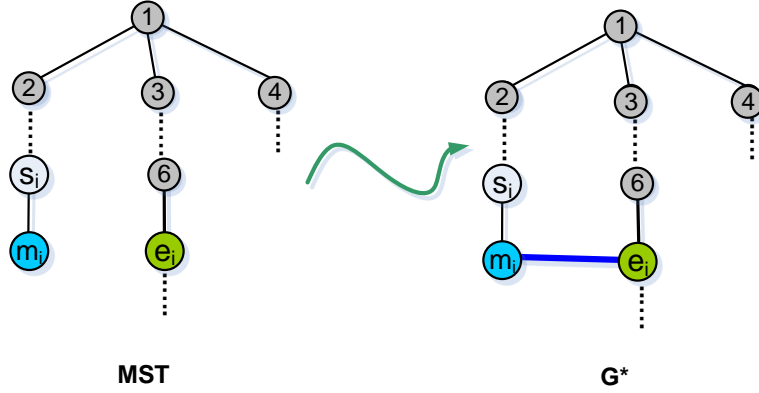


Figure 4: Step 2 of CSPP, constructing graph G^* from MST.

a two-degree node in the MST connected to s_i and e_i or a leaf node in the MST connected to s_i or e_i . In this step, for each middle node as a leaf such as m_i (suppose s_i is the parent of m_i), the edge $m_i e_i$ is added to the MST to form a graph denoted by G^* . In G^* , the degree of any middle node is even. Fig 4 shows how this step works. As a natural consequence of this step, the overall weight of the MST increases by the sum of $w(m_i, e_i)$ for any i such that m_i is a leaf middle node.

Step 3: Minimum Perfect Matching over Odd-degree Nodes

This step, which is identical to one of the Christofides' steps [6], performs *minimum perfect matching* in G'' between all odd-degree nodes in G^* ⁸ and adds the edges involved in perfect matching to G^* to build a graph denoted by \hat{G} . Since there is no odd-degree middle node in G^* , no middle node is involved in perfect matching. As a result, no infinity weight edge is added to G^* . Thus \hat{G} is still without any infinity weight edges.

Step 4: Finding an Eulerian Tour Over \hat{G}

The final output graph of Step 3, i.e., \hat{G} , is an Eulerian graph⁹. Thus, it has an Eulerian tour. An Eulerian tour is a sequence of nodes, which visits every edge exactly once. The current step (similar to Christofides') finds an Eulerian tour of \hat{G} . We call the output of this step *trail*.

Step 5: Confined Shortcut on *trail*

As defined above, *trail* is an Eulerian tour. It is possible for an Eulerian tour to visit some nodes more than once. In order to turn an Eulerian tour into a Hamiltonian tour, the extra occurrences of nodes have to be removed. An operation, called *shortcut*, is used in Christofides' algorithm [6] to do such a transformation. However, G'' contains some infinity weight edges and so it is not feasible to do shortcuts like in

⁸Number of these nodes is even because the total degrees of nodes in a graph is even.

⁹a connected graph with even-degree nodes

the Christofides' algorithm. This is because it may add infinity edges to the tour. To address this problem, we introduce a new operation, called *confined shortcut*. The procedure of this operation is discussed in the following.

Consider a node v such that it is visited more than once in *trail*. Let w and u denote the predecessor and successor in one of v 's occurrences in the tour, respectively. In this case, it is feasible to add an edge uw to the tour and remove the edges uv and vw to decrease the number of occurrences of v by one. We keep doing this process until the number of occurrences of v in the tour get to one. If one of the nodes u and w is a middle node, then the weight of the edge uw is infinity and so performing the confined shortcut operation would result in adding an infinity edge to the tour. To resolve this problem, we avoid performing the confined shortcut operation over $u-v-w$ and in place of that the operation is done over other occurrences of v . Lemma 1 shows that performing confined shortcut over at most one of the occurrences of v in *trail* may lead to adding an infinity edge. Thus, in order to build a finite Hamiltonian tour, confined shortcuts can be performed over other occurrences to avoid adding infinity weight edges. In fact, differences between shortcut in [6] and confined shortcut are in the two following issues: 1) Performing confined shortcuts in only three consecutive nodes in the tour; 2) avoiding performing confined shortcut whenever it leads to adding an infinity weight edge.

In step 5, it is possible to have a sequence like $x-p_0-p_1-\dots-p_m-y$ in which, for all i ($0 \leq i \leq m$), p_i is already visited. In Christofides' shortcut, all nodes p_i are removed in one step and then edge $x-y$ is added to the tour. On the other hand, in each step of confined shortcut, only one node is removed. Lemma 2 shows that doing confined shortcut on such cases is doable.

Lemma 1 *For each node v in trail (the Eulerian tour generated by step 4 of the CSPP algorithm), doing confined shortcut (step 5 of the algorithm) adds infinity weight edges for at most one of the v 's occurrences.* ■

Proof: According to the steps 2, 3, and 4, \hat{G} does not have any infinity edges. The node v in \hat{G} can be either head or tail of a subpath (i.e. s_i or e_i for some i) or a middle node (i.e., m_i for some i). In the former case, v (equal to s_i or e_i) is not a neighbour of a middle node in \hat{G} , unless the neighbour middle node is m_i (if v is a neighbour of m_j where $j \neq i$, then the edge vm_j has the infinity weight in \hat{G}). In the latter case, v is a middle node, say m_i . In this case, v cannot have any other middle nodes, say m_j for $j \neq i$, as its neighbour in \hat{G} , since the weight of m_im_j is infinity. Therefore, only one neighbour of v in \hat{G} would be a middle node.

Moreover, according to operations involved in steps 1 to 4, no middle node has a *multiple edge* in \hat{G} ¹⁰. As a result, at most one occurrence of v can appear next to a

¹⁰A node x has multiple edge in a graph, if there exists another node y such that there are two or more edges between x and y .

middle node in trail and doing shortcut over that specific occurrence leads to adding an infinity weight edge. \square

In \hat{G} no middle node has a multiple edge. It is worth pointing out that if \hat{G} includes a multiple edge connected to a middle node, then there exist some cases like one shown in Fig. 5, where *trail* over graph includes sequences like $s_i m_i s_i$. In this case, removing each of the occurrences of s_i (by using confined shortcut) leads to adding an infinity weight edge. Therefore, the step 5 of CSPP is not feasible in such cases, since it may add infinity weight edges to the Hamiltonian tour of CSPP.

Lemma 2 *Consider an Eulerian tour with a subsequence $x-p_0-p_1-\dots-p_m-y$ in which each node p_i , for any i ($0 \leq i \leq m$), has been already visited. Multiple applying of confined shortcut (according to step 5) on this Eulerian tour makes the number of occurrences of nodes p_i (for all i) to reduced to 1 and also it does not add any infinity edge to the tour.* \blacksquare

Proof: Since each p_i has been visited already before x in the Eulerian tour, the number of its occurrences in the tour would be more than 1. The degree of each middle node in \hat{G} is 2. Therefore, the number of occurrences of middle nodes in an Eulerian tour of this graph would be 1. This implies that none of the nodes p_i ($1 \leq i \leq m$) is a middle node (note the premise in the lemma stating that p_i has been already visited). Thus, only one of the following cases would be possible:

1. x is a middle node and y is not:

In this case, confined shortcut acts in the way as follows. First, the sequence with the first three nodes, i.e., $x-p_0-p_1$, is investigated. Since x is a middle node, removing p_0 results in adding the edge $x-p_1$, which has infinity weight. Thus, confined shortcut is not applicable on this sequence, which means p_0 is not removed. In the next sequence, i.e., $p_0-p_1-p_2$, confined shortcut is applicable, since the nodes p_0 and p_1 are not middle nodes. Applying confined shortcut on this sequence, p_1 is removed, the edge p_0-p_2 is added into the tour, and the subsequence would be turned into $x-p_0-p_2-\dots-p_m-y$. Similarly, in the next step, confined shortcut is applied on $p_0-p_2-p_3$ and p_2 is removed. This procedure is continued until the subsequence is turned into $x-p_0-y$. Since, according to Lemma 1, applying confined shortcut on at most one of the p_0 's occurrences makes adding an infinity edge (in this case, $x-p_0-p_1$), confined shortcut can be applied on other occurrences of p_0 to make the number of its occurrences equal to 1.

2. y is a middle node and x is not:

Like the case 1, multiple applying of confined shortcut, the subsequence would be turned into $x-p_m-y$. Since y is a middle node, applying confined shortcut

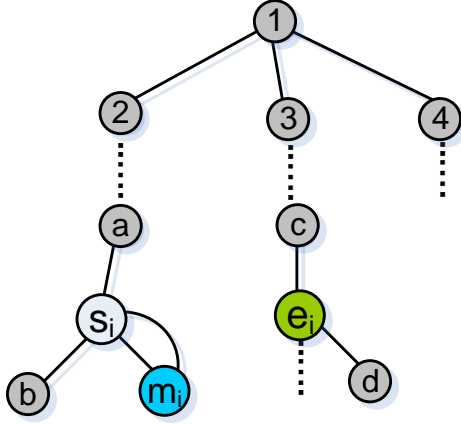


Figure 5: If a middle node has a multiple edge then step 5 of CSPP may add infinite edge to the tour.

to remove p_m would be impossible. However, according to Lemma 1, confined shortcut can be applied on other occurrences of p_m to make the number of its occurrences equal to 1.

3. Both x and y are middle nodes:

In this case, both p_0 and p_m would be problematic, i.e., applying confined shortcut on the tour does not remove p_0 and p_m . In other words, the subsequence eventually would be turned into $x-p_0-p_m-y$. However, again according to Lemma 1, this means that other occurrences of both p_0 and p_m would be removed by applying confined shortcut, which makes their number of occurrences to reduced to 1.

4. None of x and y is middle node:

In this case, all nodes p_0, \dots, p_m are removed by multiple applying of confined shortcut. Then the tour would be turned into $x-y$.

As we saw above, in all cases, the number of occurrences of each of the nodes p_0, \dots, p_m is reduced to 1 by multiple applying of confined shortcut without adding any infinity edge. \square

After the fifth step, according to Lemma 1, the resulting Hamiltonian tour, denoted by h -trail, does not include any infinity weight edge.

5.2 Ratio Bound of CSPP Algorithm

In this section, it is shown that the ratio bound of CSPP is 2. Lemmas 3 and 5 are adapted from [6]. The proofs of them can be found in Appendix. A.

Lemma 3 *The weight of MST is less than the weight of T^* , where T^* is the optimal TSP tour in G'' . ■*

Remark 2 *The notation $W(H)$ denotes the sum of edge weights of a given graph H . T^* and PM^* denote the TSP tour in G'' and edges of perfect matching in step 3, respectively.*

Lemma 4 *The total added weight to MST during the second step is less than the half of the weight of T^* . ■*

Proof: During the second step, for each middle leaf node m_i (for some i) in MST, $w(e_i, m_i)$ is added to the weight of MST. Therefore, $W(E_2) \leq \sum_{i=1}^n w(e_i, m_i) = \frac{1}{2} \sum_{i=1}^n w(s_i, m_i) + w(e_i, m_i)$, where $W(E_2)$ is the total weights of added edges to MST during this step. T^* does not include any infinity weight edge. Thus, for each subpath, say i^{th} , T^* includes the edges $s_i m_i$ and $e_i m_i$, which implies $W(T^*) > \sum_{i=1}^n w(s_i, m_i) + w(e_i, m_i)$. Therefore, the inequality $W(E_2) < 0.5W(T^*)$ holds. The lemma is proven. □

Lemma 5 *PM^* 's weight is less than half of T^* 's. ■*

Theorem 4 *The ratio bound of CSPP is 2. ■*

Proof: The graph \hat{G} is composed of the edges of the MST and the edges added in steps 2 and 3 (perfect matching). According to Lemmas 3, 4 and 5,

$$\begin{aligned} W(\hat{G}) &= W(\text{MST}) + W(E_2) + W(PM^*) \\ &\leq 2W(T^*) \end{aligned} \tag{16}$$

, where \hat{G} denotes the graph generated in step 3 of CSPP, MST denotes the output tree of step 1, E_2 denotes the set of edges which are added to MST in step 2, and PM^* denotes edges of perfect matching in step 3. Thus, the length of *trail* over \hat{G} (found in step 4) is less than 2 times of the optimal TSP tour.

trail does not contain any infinity weight edges, since the edges of \hat{G} are not infinity weight edges. Moreover, based on Lemma 1, performing shortcuts during step 5 does not add any infinity edge. Accordingly, the triangles used in performing confined shortcut (for each confined shortcut two edges of tour corresponding to a triangle are replaced with the third edge of triangle) do not include any infinity edge.

Thus, based on Theorem 2, triangle inequality is not violated in these triangles. Thus, for each execution of confined shortcuts (during step 5), the length of the tour decreases. Hence, the length of the Hamiltonian tour, $h\text{-trail}$, returned by CSPP is less than the length of $trail$ and the following statement holds:

$$W(h\text{-trail}) \leq W(trail) = W(\hat{G}) \leq 2W(T^*) \quad (17)$$

Hence, the solution produced by CSPP is within 2 of optimum. Therefore, the ratio bound of CSPP is 2. \square

5.3 Pseudo Code and Complexity Analysis of CSPP

In this section, the time complexity of the approximation algorithm CSPP is discussed in term of the number of subpaths. Algorithm 3 is a pseudo code for this algorithm, which takes G'' (the output of the IETI algorithm) as input and returns a Hamiltonian tour called $h\text{-trail}$. Let the number of subpaths in the given workspace be n . Then the graph G'' has $3n$ nodes and $9n^2$ edges.

Complexity analysis of Step 1 (line 1): MST can be found using Kruskal's algorithm [20] or Prim's algorithm [31]. Two implementations of these algorithms are *Improved Implementation of Kruskal Algorithm* and *Fibonacci Heap Implementation of Prim's Algorithm* [1], respectively, which both are in $O(|E| + |V| \log |V|)$ complexity class. Since the numbers of nodes and edges are $3n$ and $9n^2$, respectively, the total complexity of this step would be in $O(n^2)$.

Complexity analysis of Step 2 (lines 2 to 9): This step includes a loop of n iterations, where each iteration takes $O(1)$ time. Thus, the total complexity of this step is in $O(n)$.

Complexity analysis of Step 3 (lines 10 and 11): The modified version of Edmond's Blossom Shrinking algorithm in [25] can be used to do perfect matching, which requires $O(|V|^3)$ running time. The total number of odd-degree nodes in G^* is in the $O(n)$ class. Thus, complexity of perfect matching would be in the $O(n^3)$ class. Furthermore, the number of nodes involved in perfect matching is in $O(n)$. Therefore, adding the perfect matching edges to G^* (line 11) is in $O(n)$. As a result, step 3 requires $O(n^3)$ running time.

Complexity of Step 4 (line 12): One can use Fleury algorithm [29] to find Eulerian tour, which has the running time complexity of $O(|E|)$. \hat{G} includes the following kinds of edges: 1) the edges of MST ($O(n)$ edges) 2) the edges between $m_i s_i$ or $m_i e_i$, which are added in step 2 ($O(n)$ edges) 3) the edges of perfect matching ($O(n)$ edges). Thus, the total number of the edges in \hat{G} is in the $O(n)$ class. This implies that the running time of Fleury algorithm on \hat{G} is in the $O(n)$ class.

Complexity analysis of Step 5 (lines 13 to 34): $trail$ in line 12 is the output of step 4. In line 13, $trail$ is copied into $h\text{-trail}$. Hence, according to Lemma 1, confined shortcut (step 5) can be applied on $h\text{-trail}$ in order to produce a finite Hamiltonian

Algorithm 3 CSPP

```
1: Find the minimum spanning tree, MST
2: for all subpaths such as  $i$  do
3:   if  $m_i$  is a leaf node and  $s_i$  is the parent of  $m_i$  then
4:     Add edge  $m_i e_i$ 
5:   end if
6:   if  $m_i$  is a leaf node and  $e_i$  is the parent of  $m_i$  then
7:     Add edge  $m_i s_i$ 
8:   end if
9: end for
10: Do perfect matching over odd-degree nodes in  $G^*$ 
11: Add perfect matching edges to  $G^*$  and construct  $\hat{G}$ 
12: Find Eulerian tour,  $trail$ , in  $\hat{G}$ 
13:  $h\text{-}trail \leftarrow trail$ 
14: Make  $visit$  zero for all nodes in  $G''$ 
15: while has not reached the end of  $h\text{-}trail$  do
16:   select the next node in  $t$  in forward order
17:    $visit(t) \leftarrow visit(t) + 1$ 
18:   if  $visit(t) \geq 2$  then
19:      $t_0 \leftarrow$  the node appears before  $t$  in  $h\text{-}trail$ 
20:      $t_1 \leftarrow$  the node appears after  $t$  in  $h\text{-}trail$ 
21:     if  $w(t_0, t_1)$  is finite then
22:       Delete node  $t$ 
23:        $visit(t) \leftarrow visit(t) - 1$ 
24:     end if
25:   end if
26: end while
27: while has not reached the end of  $h\text{-}trail$  do
28:   Select the next node  $t$  in  $h\text{-}trail$  forward order
29:   if  $visit(t) == 2$  then
30:     Delete node  $t$ 
31:      $visit(t) \leftarrow visit(t) - 1$ 
32:   end if
33: end while
34: return  $h\text{-}trail$ 
```

tour. The procedure of confined shortcut is shown in lines 14 to 33. In order to perform confined shortcut, $h\text{-trail}$ is explored two times from head to tail (lines 15 to 26 and lines 27 to 33). If during an exploration (moving along all members of $h\text{-trail}$) performing a confined shortcut leads to adding an edge with infinity weight, that particular shortcut will not be performed.

One can think of $h\text{-trail}$ as a list of nodes. Lines 16 to 25 iterates over all members of this list. At each iteration, $visit(t)$ shows the number of occurrences of node t in $h\text{-trail}$ up to the current index of the list. If the number of occurrences of a node is more than 1 (line 18), its current occurrence is removed (line 22) provided that performing confined shortcut (t_0-t-t_1) does not add an infinity edge (line 21). As a result, only the first occurrence of a node is kept and the rest are removed. According to Lemma 1, two scenarios can happen for a node t in $h\text{-trail}$: 1) performing confined shortcut over one of its occurrences adds an infinity edge, 2) performing confined shortcut over each occurrence never adds an infinity edge. Therefore, after executing the loop in lines 15 to 26, one of the three following cases can occur for a node t :

1. Performing confined shortcut over each occurrence of t does not add an infinity edge. In this case, only the first occurrence of t remains in $h\text{-trail}$ and the others are removed. This makes the value of $visit(t)$ equal to 1.
2. Performing confined shortcut over the first occurrence of t adds an infinity edge. Similar to the previous case, only the first occurrence is kept in the list and the others are removed. Therefore, the value of $visit(t)$ would be 1 after executing the loop.
3. Performing confined shortcut over one of the t 's occurrences (not the first one) adds an infinity edge. In this case, two occurrences are kept and the others are removed. The kept occurrences are the first one and the occurrence that results in adding an infinity edge. In this case, the value of $visit(t)$ would be 2.

As already mentioned, to turn an Eulerian tour to a Hamiltonian one, the occurrences of each node must be reduced to one. To this end, the execution of lines 27 to 33 reduces the value of $visit$ for those nodes that are visited more than once (nodes in the case 3 above).

Performing confined shortcut on the second occurrence of nodes belonging to the case 3 leads to adding an infinity edge. Therefore, confined shortcut must be performed over the first occurrence of nodes. Lines 28 to 32 iterate over each element of $h\text{-trail}$. Line 29 detects the first occurrence of nodes belonging to the case 3. Then, confined shortcut is performed on the first occurrence and the number of visits is reduced to one (line 30).

At the end of the first exploring, the number of occurrences of each node can be up to two, while at the end of the second exploring, this number is exactly one. These explorations can prevent us from moving constantly back and forth over $h\text{-trail}$.

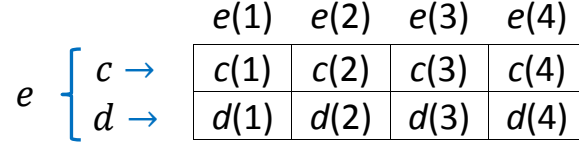


Figure 6: An augmented chromosome used in Gyorfi_GA which consists of two arrays (adapted from [16]).

The time of exploring *h-trail* is linearly dependent to the number of its edges, which is equal to the number of \hat{G} 's edges (the number of edges in \hat{G} is in the $O(n)$ class). As a consequence, the complexity of each exploration is in $O(n)$ and, hence, the overall complexity of this step would be in $O(n)$.

According to the complexity class of each step (discussed above), the total complexity of CSPP would be in the $O(n^3)$ class.

6 Experiments and Results

In this section, we empirically gauge the performance of CSPP by comparing it with the method proposed by Gyorfi *et al.* [16], which uses GA. For simplicity, throughout this section, this method is referred by Gyorfi_GA. For comparison, we use different sets of workspaces including 3 workspaces with 20 subpaths, 3 workspaces with 50 subpaths and 3 workspaces with 80 subpaths. The subpaths of each workspace were built randomly in different lengths and different locations. Any of the 9 randomly generated workspaces can be seen as a sample for a real work application. For instance, they can be a scratch on a surface that a robot should smooth.

In the Gyorfi_GA method, each feasible solution of the problem is shown with a fixed length augmented chromosome with length of n . Each augmented chromosome such as e , consists two chromosomes such as c and d shown in Fig. 6. Each gene of chromosome d can take value 0 or 1, which indicates the connection between two subpath. For any i , $d(i) = 0$ shows that the head of the subpath $c(i)$ is connected to the subpath $c(i + 1)$. Similarly, $d(i) = 1$ shows that the tail of $c(i)$ is connected to $c(i + 1)$. Five different genetic operations are used in the Gyorfi_GA method. “The crossover operator produces an offspring chromosome by combining genes from two parent chromosomes. The inversion operator changes a region of a parent chromosome by inverting the order of the genes in the region. The rotation operator changes a region of a parent chromosome by rotating the genes in the region in manner similar to a circular shift register. The mutation operator exchanges two genes in a parent chromosome. The subpath reversal operator changes the direction flag of randomly chosen genes within a parent chromosome”[16].

Parameter setting for the operation rates of Gyorfi_GA were chosen to be 0.5 for

crossover, 0.25 for inversion, 0.25 for rotation, 0.5 for mutation and 0.5 for subpath reversal. These parameters are fixed during all experiments and seem to be near optimal set of parameters, according to the results of different parameter settings. Each of the methods were executed 30 times over each environment.

Table 1 compares means and standard deviations (in parenthesis) of 30 executions of CSPP and Gyorfi_GA (population size is 100) in three environments with 20 subpaths. Gyorfi_GA iterates until it converges (all of the executions converged at 150 iterations at most). The results indicate that Gyorfi_GA returns longer outputs with more time consumption (in average CSPP is more than 8.55 times efficient than Gyorfi_GA in terms of time consumption).

Likewise, both methods were executed in the workspaces with 50 and 80 subpaths over 30 iterations the results of which are shown in Tables 2 and 3. The population sizes for these experiments are 200 and 300, respectively. Roughly, the Gyorfi_GA converges after 300 and 500 iterations for workspaces with 50 and 80 subpaths. According to Tables 2 and 3, Gyorfi_GA leads to less efficient results both in terms of length of output and time. Technically speaking, the time that the Gyorfi_GA method spends is more than 7.28 and 2.97 times than CSPP uses to generate the final results.

According to Table 4, using CSPP in place of Gyorfi_GA provides an average improvement of 86.7% and 0.33% in execution time and result's length, respectively. Note that these improvements happen while the average rate of deviation in both execution time and result's length in Gyorfi_GA (0.13 and 32.2, respectively) are reduced to 0 in CSPP.

Similarly, according to Tables 5 and 6, CSPP improves the average execution time

Table 1: Results of CSPP and Gyorfi_GA with population size 100, in three workspaces with 20 subpaths.

Env.	CSPP		Gyorfi_GA	
	Length	Time	Length	Time (in seconds)
1st	1134.6(0)	0.6(0)	1147.9(31.7)	5.6(0.2)
2nd	1375.8(0)	1(0)	1357.6(34.5)	3.9(0.1)
3rd	1394.8(0)	0.2(0)	1410.8(30.3)	5.9(0.1)

Table 2: Results of CSPP and Gyorfi_GA with population size 200, in three workspaces with 50 subpaths.

Env.	CSPP		Gyorfi_GA	
	Length	Time	Length	Time (in seconds)
1st	2889.8(0)	6.0(0)	3052.9(49.3)	65.1(3.1)
2nd	3004.5(0)	6.3(0)	3110.7(45)	64.2(5.1)
3rd	2710.1(0)	12.7(0)	2843.0(54)	52.1(3)

Table 3: Results of CSPP and Gyorfi_GA with population size 300, in three workspaces with 80 subpaths.

Env.	CSPP		Gyorfi_GA	
	Length	Time	Length	Time (in seconds)
1st	4536.2(0)	74.8(0)	4791.4(71.5)	168.2(19.7)
2nd	4254.2(0)	16.7(0)	4610.2(81.3)	147.8(2.8)
3rd	4732.3(0)	65.5(0)	5003.5(78.1)	149.7(3.3)

Table 4: Result Improving by CSPP in comparison with Gyorfi_GA in environments with 20 subpaths.

Env.	Time Improving	Length Improving
I	89.3%	1%
II	74.4%	-1%
III	96.6%	1%
Average	86.7%	0.33%

85.5% and 66.8% in workspaces with 50 and 80 subpaths, respectively. Also, result's length is improved 4.43% and 6.13% in average, respectively. The average rate of deviation in both execution time and result's length in Gyorfi_GA are reduced to 0 in CSPP. Note that the average rate of deviation in time and length in Gyorfi_GA are, respectively, 3.73 and 49.4 in workspaces with 50 subpaths and 8.6, 77 in workspaces with 80 subpaths.

Comparing the outputs of the two algorithm, we get the following results:

1. Increasing the number of subpaths causes a pretty huge increase in the rate of deviation in both execution time and result's length in Gyorfi_GA. This implies that uncertainty in Gyorfi_GA increases by increasing the number of subpaths in the same environment, whereas this is not the case in CSPP. Since the rate of deviation is 0 for both execution time and result's length in CSPP, it always delivers a certain result in a certain time for a given input.
2. As we see in the third columns of the Tables 4, 5, and 6, the CSPP algorithm delivers better results than Gyorfi_GA does. Also, it is more intriguing that the efficiency of CSPP remarkably increases in comparison with Gyorfi_GA by increasing the number of subpaths: the average length improvement increases from 0.33% to 6.13% by increasing the number of subpaths from 20 to 80.
3. The CSPP algorithm is much more faster than the Gyorfi_GA algorithm to solve SPP in all above experiments.

Table 5: Result Improving by CSPP in comparison with Gyorfi_GA in environments with 50 subpaths.

Env.	Time Improving	Length Improving
I	90.7%	5.3%
II	90.2%	3.4%
III	75.6%	4.6%
Average	85.5%	4.43%

Table 6: Result Improving by CSPP in comparison with Gyorfi_GA in environments with 80 subpaths.

Env.	Time Improving	Length Improving
I	55.5%	5.3%
II	88.7%	7.7%
III	56.2%	5.4%
Average	66.8%	6.13%

7 Conclusion and Future Works

In this paper, we have proposed a method to transform SPP to TSP. Transforming SPP to TSP allows us to use existing algorithms for solving TSP in the subpath planning context. However, violation of triangle inequality hinders applying existing fixed-ratio bound approximation algorithms such as Christofides for SPP. To address this problem, we have proposed the algorithm IETI, which makes a main subset of violating triangles to satisfy the triangle inequality condition. The IETI algorithm should be seen as a fundamental step in proposing and applying fixed-ratio bound approximation algorithms for solving SPP. Using this method, a fixed-ratio bound approximation algorithm, called CSPP, has been proposed to find a near optimal solution for SPP. CSPP Solves TSP over the output graph of IETI algorithm. CSPP is similar to the Christofides’ algorithm regarding time complexity, but its ratio bound (the fixed ratio bound of 2) is more than the Christofides’ algorithm. This is natural, since CSPP can be executed over some graphs, which cannot be in the domain of Christofides’ algorithm.

Although we theoretically showed that CSPP has reasonable time complexity with a small fixed ratio bound, experiments also indicate that the algorithm is a fast algorithm with more efficient results in comparison with a state-of-the-art method Gyorfi_GA. Moreover, the differences between efficiency of these algorithms becomes much more significant as the number of subpaths increases.

We think that it would be possible to improve the results of the CSPP method using some improvement methods such as Lin-Kernighan [23] and Helsgaun [17].

We plan to propose some algorithms with smaller ratio bounds. Indeed, we try to improve the ratio bound of the CSPP algorithm using some heuristic-based methods. One important guide line could be using the graph generated by the IETI algorithm as an input for a modified version of the RPP algorithm [11].

Our algorithms solve SPP without considering any constraints, say some priorities over subpaths or environments with obstacles. We also plan to solve SPP for some certain applications with some constraints.

CSPP is neither a meta-heuristic nor a stochastic method. It may be possible to propose some meta-heuristic based methods such as tabu search and simulated annealing based on the TSP model of a given SPP to generate more efficient results in the case of offline tasks. We can initialize such algorithms with the solutions found by CSPP to improve the output of such meta-heuristic methods.

References

- [1] R. Ahuja, T. Magnanti, and J. Orlin. Network flows: Theory, algorithms, and applications, 1993.
- [2] E. M. Arkin and R. Hassin. Approximation algorithms for the geometric covering salesman problem. *Discrete Applied Mathematics*, 55(3):197–218, 1994.
- [3] G. Ausiello. *Complexity and approximation: Combinatorial optimization problems and their approximability properties*. Springer Verlag, 1999.
- [4] C. Blum and A. Roli. Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys (CSUR)*, 35(3):268–308, 2003.
- [5] D. Chen, S. Luan, and C. Wang. Coupled path planning, region optimization, and applications in intensity-modulated radiation therapy. *Algorithmica*, 60(1):152—174, 2011.
- [6] N. Christofides. Worst-Case analysis of a new heuristic for the travelling salesman problem. Technical report, DTIC Document, 1976.
- [7] A. Coja-Oghlan, S. O. Krumke, and T. Nierhoff. A heuristic for the stacker crane problem on trees which is almost surely exact. *Journal of Algorithms*, 61(1):1–19, 2006.
- [8] A. Dumitrescu and J. S. Mitchell. Approximation algorithms for tsp with neighborhoods in the plane. In *Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms*, pages 38–46. Society for Industrial and Applied Mathematics, 2001.
- [9] H. A. Eiselt, M. Gendreau, and G. Laporte. Arc routing problems, part i: The chinese postman problem. *Operations Research*, 43(2):231–242, 1995.
- [10] H. A. Eiselt, M. Gendreau, and G. Laporte. Arc routing problems, part ii: The rural postman problem. *Operations Research*, 43(3):399–414, 1995.
- [11] G. N. Frederickson. Approximation algorithms for some postman problems. *Journal of the ACM (JACM)*, 26(3):538–554, 1979.
- [12] G. N. Frederickson, M. S. Hecht, and C. E. Kim. Approximation algorithms for some routing problems. In *Foundations of Computer Science, 1976., 17th Annual Symposium on*, pages 216–227. IEEE, 1976.

- [13] M. Garey and D. Johnson. Computers and intractability. a guide to the theory of np-completeness. a series of books in the mathematical sciences, ed. v. klee. 1979.
- [14] R. Geraerts and M. Overmars. The corridor map method: a general framework for real-time high-quality path planning. *Computer Animation and Virtual Worlds*, 18(2):107—119, 2007.
- [15] D. Goldberg. *Genetic algorithms in search, optimization, and machine learning*. Addison-wesley, 1989.
- [16] J. Györfi, D. Gamota, S. Mok, J. Szczech, M. Toloo, and J. Zhang. Evolutionary path planning with subpath constraints. *IEEE Transactions on Electronics Packaging Manufacturing*, 33(2):143—151, 2010.
- [17] K. Helsgaun. An effective implementation of the Lin-Kernighan traveling salesman heuristic. *European Journal of Operational Research*, 126(1):106—130, 2000.
- [18] D. Johnson and L. McGeoch. The traveling salesman problem: A case study in local optimization. *Local search in combinatorial optimization*, pages 215—310, 1997.
- [19] R. Karp. *Reducibility among combinatorial problems*, volume 40. Springer, 1972.
- [20] J. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical society*, 7(1):48—50, 1956.
- [21] G. Laporte. The traveling salesman problem: An overview of exact and approximate algorithms. *European Journal of Operational Research*, 59(2):231—247, 1992.
- [22] J. K. Lenstra and A. Kan. Complexity of vehicle routing and scheduling problems. *Networks*, 11(2):221—227, 1981.
- [23] S. Lin and B. Kernighan. An effective heuristic algorithm for the traveling-salesman problem. *Operations research*, pages 498—516, 1973.
- [24] C. S. Mata and J. S. Mitchell. Approximation algorithms for geometric tour and network design problems. In *Proceedings of the eleventh annual symposium on Computational geometry*, pages 360—369. ACM, 1995.
- [25] S. Micali and V. Vazirani. An $O((V|V|C|E|))$ algorithm for finding maximum matching in general graphs. In *Foundations of Computer Science, 1980., 21st Annual Symposium on*, pages 17—27, 1980.

- [26] R. G. Michael and D. S. Johnson. Computers and intractability: A guide to the theory of np-completeness. *WH Freeman & Co., San Francisco*, 1979.
- [27] C. Orloff. A fundamental problem in vehicle routing. *Networks*, 4(1):35–64, 1974.
- [28] C. H. Papadimitriou. The euclidean travelling salesman problem is np-complete. *Theoretical Computer Science*, 4(3):237–244, 1977.
- [29] S. Pemmaraju and S. Skiena. *Computational discrete mathematics: combinatorics and graph theory with Mathematica*. Cambridge Univ Pr, 2003.
- [30] R. Pepy, M. Kieffer, and E. Walter. Reliable robust path planning with application to mobile robots. *International Journal of Applied Mathematics and Computer Science*, 19(3):413–424, 2009.
- [31] R. Prim. Shortest connection networks and some generalizations. *Bell system technical journal*, 36(6):1389–1401, 1957.
- [32] S. Safra and O. Schwartz. On the complexity of approximating tsp with neighborhoods and related problems. *computational complexity*, 14(4):281–307, 2006.
- [33] S. Sahni and T. Gonzalez. P-complete approximation problems. *Journal of the ACM (JACM)*, 23(3):555–565, 1976.
- [34] W. Sheng, H. Chen, N. Xi, and Y. Chen. Tool path planning for compound surfaces in spray forming processes. *Automation Science and Engineering, IEEE Transactions on*, 2(3):240–249, 2005.
- [35] G. Song, S. Thomas, K. DILL, J. Scholtz, and N. Amato. A path planning-based study of protein folding with a case study of hairpin formation in protein g and l. In *Proc. Pacific Symposium of Biocomputing (PSB)*, pages 240–251, 2002.
- [36] L. Tapia, S. Thomas, and N. Amato. A motion planning approach to studying molecular motions. *Communications in Information & Systems*, 10(1):53–68, 2010.
- [37] G. Tong-ying, Q. Dao-kui, and D. Zai-li. Research of path planning for polishing robot based on improved genetic algorithm. In *Robotics and Biomimetics, 2004. ROBIO 2004. IEEE International Conference on*, pages 334–338, 2004.
- [38] K. Treleaven, M. Pavone, and E. Frazzoli. Asymptotically optimal algorithms for one-to-one pickup and delivery problems with applications to transportation systems. 2013.

A Proofs of Lemma 3 and Lemma 5

The proofs of Lemmas 3 and 5 are adapted from [6].

Proof: [Proof of Lemma 3] Suppose H is a Hamiltonian tour which is constructed by subtracting an arbitrary edge of T^* . Note that H is a spanning tree of G . Hence:

$$W(MST) \leq W(H) < W(T^*) \quad (18)$$

□

Proof: [Proof of Lemma 5] For an n -city TSP, consider $T^* = (x_{i_1}, x_{i_2}, \dots, x_{i_n})$. Starting from vertex x_{i_1} and travelling round the circuit T^* , allocate the links traversed in an alternating manner to two sets M_1 and M_2 . Starting with M_1 , for example: $M_1 = \{(x_{i_1}, x_{i_2}), (x_{i_3}, x_{i_4}), \dots, (x_{i_{n-1}}, x_{i_n})\}$ and $M_2 = \{(x_{i_2}, x_{i_3}), (x_{i_4}, x_{i_5}), \dots, (x_{i_n}, x_{i_1})\}$. M_1 and M_2 are matching of G and $W(M_1) + W(M_2) = W(T^*)$. Since M_1 and M_2 are defined arbitrarily we can assume $W(M_1) \leq W(M_2)$ without loss of generality, and so we have: $W(PM) \leq W(M_1) \leq \frac{1}{2}W(T^*)$. Hence, the lemma is proved. □